

ABSTRACT

The present invention is directed to a network switch that determines when specific content is hot and directs flow to one or more cache servers. The architecture can include a tag generator to generate unique tags corresponding to a server in a plurality of servers, a content pre-fetching algorithm to retrieve information before the information is requested by determining the hotness of the information, and a cache server in which stored information is configured based upon the relative degrees of hotness of the stored information.

M:\4366\50\Patent Application.wpd